



# OPEN DATA

---

1. Définition, enjeux
2. Les données, gisements
3. Licences, formats, outils d'exploitation

# Open data à la loupe



Spot de présentation de la démarche d'ouverture des données numériques publiques, initié par LiberTIC, soutenu par Nantes Métropole et réalisé par A2B Production en licence Art Libre.

# Open Data c'est quoi?

- **Définition** : L'*open data* ou **donnée ouverte** est une donnée numérique dont l'accès et l'usage sont laissés libres aux usagers. Elle peut être **d'origine publique ou privée**, produite notamment par une collectivité, un service public ou une entreprise. Elle est diffusée de manière structurée selon une méthode et une **licence ouverte** garantissant son libre accès et sa réutilisation par tous, sans restriction technique, juridique ou financière.
- Ces droits d'accès et de réutilisation s'inscrivent dans la pensée qui considère l'information publique comme un bien commun dont la diffusion est d'intérêt public et général.
- L'ouverture des données (*open data*) est à la fois **un mouvement, une philosophie d'accès à l'information** et une **pratique de publication de données** librement accessibles et exploitables.
- Le mouvement s'est étendu notamment sous l'impulsion d'ONG comme l'[Open Knowledge Foundation](#) (OKFN) et le [Partenariat pour un gouvernement ouvert](#) (PGO).

# *Open knowledge foundation*

- L'**Open Knowledge Foundation** est une association à but non lucratif de droit britannique promouvant la culture libre, en particulier les contenus libres et l'*open data* (données ouvertes). Elle a été créée le 24 mai 2004 à Cambridge.
- 3 groupes de travail :
  - [Open research](#) groupe de travail dédié au domaine de la recherche ouverte et mettant l'accent sur le degré d'ouverture des données et sur comment Internet et les technologies numériques peuvent soutenir de nouvelles formes de collaboration et de partage dans la recherche scientifique.
  - [Open culture](#) groupe de travail voués à l'étude des moyens par lesquels des logiciels open source et les contenus culturels ouverts peuvent accroître l'accès à notre patrimoine culturel et forger de nouveaux processus de création et de collaboration.
  - [Open government](#) il s'agit de groupes de travail qui étudient le domaine des données publiques ouvertes, et qui sont tous engagés à construire des communautés qui rendront les **gouvernements plus transparents et responsables**.

# Une démarche openGov

## Comment améliorer la démocratie ?

Les valeurs d'une démocratie ouverte



Schéma démocratie ouverte de Arnel Le Coz et Cyril Lage  
est mis à disposition selon les termes de la licence Creative Commons Attribution

# Une démarche openGov

<https://www.democratieouverte.org>

- **donner accès** à une meilleure connaissance du territoire : espace public et naturel, patrimoine, équipements, voirie, cartographie, information socio-démographique, etc.
- **renforcer la transparence** de l'action publique : budgets, dépenses, subventions, coûts des prestation, sécurité publique, etc.
- **développer les services** et de soutenir l'économie locale, améliorer la qualité des services rendus aux publics, participer aux actions en faveur du développement durable et par une dynamique territoriale, renforcer l'attractivité du territoire.
- **Mutualisation, amélioration** et efficacité des échanges entre services

Exemple : le projet [Kawaa](#)

# Quelles données ?

Toutes les données structurées et produites par des acteurs publics ou privés dans le cadre d'une mission de service public. Cette définition s'étend aux données décrivant l'espace, les services publics et l'exercice politique, produites par des acteurs non institutionnels : associations, citoyens engagés, acteurs économiques ou académiques.

## **Sont exclus de l'opendata**

### • Les données à caractère personnel

« Constitue une donnée à caractère personnel toute information relative à une personne physique identifiée directement ou indirectement par référence à un ou plusieurs éléments qui lui sont propres » (la loi 78-17 du 6 janvier 1978 art 2)

Ex : nom, adresse, photographie, num. tél., num. d'identifiant, adresse IP, adresse électronique...

### • les documents administratifs dont la consultation ou la communication porterait atteinte :

A-6a) Au secret des délibérations du Gouvernement et des autorités responsables relevant du pouvoir exécutif ;

A-6d) A la sûreté de l'Etat, à la sécurité publique ou à la sécurité des personnes;

A-10c) Ou sur lesquels des tiers détiennent des droits de propriété intellectuelle.



# Données publiques versus Big Data

## **Données publiques : services publics, intérêt général, ouvertes**

- Effectifs: d'agents municipaux, d'élèves
- Horaires : bus et tram en temps réel,
- Patrimoine : arbres, équipements sportifs, points lumineux, bornes fontaines ...
- Éléments liés à l'exploitation : livres empruntés dans les bibliothèques, budget et compte administratif, marchés publics
- Description du territoire : photo aérienne, altimétrie, description des rues,
- Reflet de la vie locale : mariage, naissance, prénoms, élections, agendas...

## **Big data: données personnelles, comportement individuel, données privées**

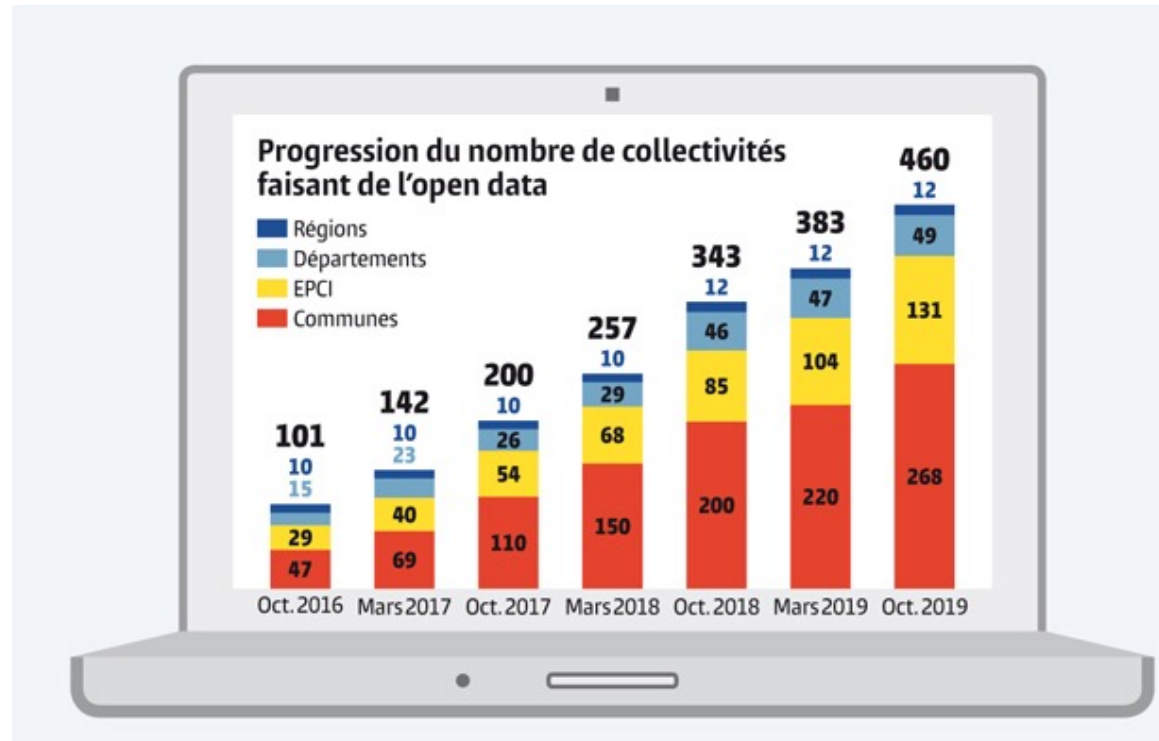
- Traces (téléphone mobile, connexion)
- Navigation internet, réseaux sociaux, messagerie,
- Achats (CB), déplacements individuels ou collectifs (badge, carte)
- Consommations individuelles (eau, électricité, ..)
- Objets connectés

# Gisement de données

- **Collectivités**

- Accompagnement pédagogique :  
[DataLab](#), [InfoLab](#)
- [OpenDataSoft](#)

[Youtube : Collectivités et open data : comment, pourquoi, pour qui ?](#)



Source : [la gazette des communes](#)

- **Etat**

- [EtatLab : Missions](#)
- [data.gouv.fr](#)
  - 39 316 jeux de données ouvertes
  - 2792 applications ré-utilisatrices

# Gisement de données

- **Citoyen**

- [Regards Citoyens](#),
- [DataPublica](#),
- [Data4Citizen](#)

- **Instituts**

- [INSEE](#), [INPI](#), [LEGIFRANCE](#)
- [IDG \(infrastructure de données géographiques\)](#)

- **Agences de l'air, de l'eau**

- [Laboratoire Central de Surveillance de la Qualité de l'Air](#), [Carte mondiale](#), [IQAir](#),
- [Données publiques sur l'eau en France](#)

- **Mondial**

- [Global open data index](#) (*Open Knowledge Foundation*)
- [La Banque Mondiale](#)

# Licences

## Licences Opendata stabilisées en France autour de :

- **ODbL** : **Open Data Base Licence**, édité par *OpenKnowledge Foundation*
  - **Licence Ouverte** : édité par Etalab
- Elles autorisent la réutilisation y compris à des fins commerciales.
  - Elles imposent des conditions de Mention et d'Intégrité
  - Elles n'imposent pas d'obligation pour le producteur
  - ODbL est plus restrictive car elle impose la notion de « Partage à l'Identique »

# Licences ODbL

## Vous êtes libres :



De partager : copier, distribuer et utiliser la base de données.



De créer : produire des créations à partir de cette base de données.



D'adapter : modifier, transformer et construire à partir de cette base de données.

## Aussi longtemps que :



**Vous mentionnez la paternité :** Vous devez mentionnez la source de la base de données pour toute utilisation publique de la base de données, ou pour toute création produite à partir de la base de données, de la manière indiquée dans l'ODbL. Pour toute utilisation ou redistribution de la base de données, ou création produite à partir de cette base de données, vous devez clairement mentionner aux tiers la licence de la base de données et garder intacte toute mention légale sur la base de données originale.



**Vous partagez aux conditions identiques :** si vous utilisez publiquement une version adaptée de cette base de données, ou que vous produisiez une création à partir d'une base de données adaptée, vous devez aussi offrir cette base de données adaptée selon les termes de la licence ODbL.



**Gardez ouvert :** si vous redistribuez la base de données, ou une version modifiée de celle-ci, alors vous ne pouvez utiliser de mesure technique restreignant la création que si vous distribuez aussi une version sans ces restrictions.

# Formats

- Pour être Opendata, les données doivent être :
  - au format électronique,
  - accessible sur internet,
  - structurées,
  - à un format ouvert (et de préférence non propriétaire).
- **Types de format des fichiers disponibles**
  - Données alpha-num. : csv, ods, xls, doc, xml, JSON...
    - Un document au format pdf, même full text –et pas image- est lisible par une machine mais n'est ouvrable que par Acrobat/Adobe, donc non Opendata
  - Données géographiques : Shape, GéoJSON, KML, KMZ,
    - Format propriétaires non ouverts : DWG et DXF (Autocad)
  - Format spécifiques :
    - Domaine Transport : [GTFS](#), Chouette, [NetEX](#), Neptune...
    - GPX (ouvert), ...

# Outils d'exploitation

## Données Alphanumériques

- Suites bureautiques habituelles : OpenOffice, Libre Office,
  - Et les outils commerciaux habituels (Google, MS, ...)
- De nombreux outils en ligne : ConversionTools Services (xml<>csv...)
- Les langages de programmation habituels : C, Python...
- [OpenRefine](#) (previously Google Refine) pour l'exploration et le retraitement des données

## Données géographiques

- [QGIS](#) : vraie suite logicielle de géomatique (SIG), convivial
- [UMAP \(openstreetmap\)](#), [GeoServer](#) : manipulation des cartes
- Non OpenSource : ARCGis, Autocad, MapInfo, GéoMap, GooglePro&Co

# Exemples de réutilisations

- Handicap : [Handimap](#), [wheelmap](#)
- Train temps réel : [raildar](#), [snCF geolocalisation](#)
- Parking temps réel : [Nantes](#), [Rennes](#)
- Transport en commun multi-modal : [Toulouse](#), [Lyon](#)
- Développement logiciels libres : [makina corpus](#)



# Des sources d'info sur l'open data

## Documents

- <http://opendatahandbook.org/guide/fr/>  
(*Open Knowledge Foundation*)

## Youtube

- [\\*\\*\\*\\* L'Open Data à la Loupe](#). Yann Bresson
- [Quel est l'impact de l'Open Data ?](#) Charles RUELLE, 2014
- [Open Data explained in a nutshell](#), Simpleshow foundation, 2016
- [\\*\\*\\*\\* Comment nous avons trouvé la pire place de parking de NY ?](#), Ben Wellington, TEDxNewYork
- [L'Open Data, Avenir des Big Data](#), Jean Marc LAZAR, TEDxUTCompiègne

# Pire place de parking de NY !



Les agences municipales ont accès à une richesse de données et de statistiques décrivant chaque aspect de la vie urbaine. Mais comme l'analyste Ben Wellington le suggère dans cette conférence amusante, elles ne savent parfois pas quoi faire avec. Il montre comment une combinaison de questions inattendues et un traitement intelligent des données peut produire des idées étrangement utiles, et partage des conseils sur comment publier ces énormes jeux de données pour que n'importe qui puisse les utiliser.

# LINKED OPEN DATA

---

1. Web des données, *Linked Data*
2. *Resource Description Framework (RDF)*
3. SparQL

# 1. LOD : Qu'est-ce que le Web ?

Un espace documentaire décentralisé, interconnecté, interopérable et *évolutif*.

- décentralisé → **HTTP 2014, HTTP 2.0**
- interconnecté → **URL, URI, IRI**
- interopérable → **HTML5, CSS, JS**

## **Vers un Web de données**

Un espace **de données** décentralisé, interconnecté et interopérable.

- décentralisé → **HTTP**
- interconnecté → **URL**
- interopérable → **?**

# 1. LOD : Web de ressources

Le web est constitué de **ressources**, par exemple :

- le bulletin météo du jour pour Lyon
- le bulletin météo du jour pour le lieu courant
- ma commande de café de jeudi dernier

Chaque ressource est identifiée par un IRI (*Internationalized Resource Identifier*), e.g.:

- <http://meteo.example.com/lyon>
- <http://meteo.example.com/ici>
- <http://commerce.example.com/commande/192837>

# 1. LOD : Web de ressources

Une ressource n'est pas un simple fichier, dont on récupérerait le contenu.

Elle est un objet *actif*, avec lequel on interagit.

- le bulletin météo du jour pour Lyon :  
→ le contenu change régulièrement
- le bulletin météo du jour pour le lieu courant :  
→ le contenu dépend de plus du contexte de l'utilisateur
- ma commande de café de jeudi dernier :  
→ on peut agir dessus (par exemple pour l'annuler)

# 1. LOD : Ressources et représentations

- Une ressource n'est jamais manipulée directement, mais toujours à travers des **représentations** (pour la créer, la consulter, la modifier).
- Les représentations d'une ressource peuvent varier en fonction
  - de son *état*
  - de l'agent qui manipule la ressource (négociation de contenu, contexte)

représentation :	utilisable par :
texte (HTML...)	humains, moteurs de recherche
médias (image, son...)	<i>surtout</i> humains
données structurées	machines

# 1. LOD : de HTML à XML

XML (*eXtensible Markup Language*) a été recommandé par le W3C en 1998. L'objectif était de pallier la sémantique « faible » de HTML.

```
<!-- HTML -->  
<a href="http://champion.net/">  
  Pierre-Antoine <strong>Champion</strong>  
  (<em>Maître de conférences</em>)</a>
```

```
<!-- XML -->  
<Person homepage="http://champion.net/">  
  <givenName>Pierre-Antoine</givenName>  
  <familyName>Champion</familyName>  
  <job>Maître de conférences</job></Person>
```



# 1. LOD : de XML à RDF

- Le modèle sous-jacent de la syntaxe XML est un arbre (*XML Infoset*), ce qui n'est pas adapté à la structure décentralisée du Web.
- L'objectif du *Resource Description Framework* (RDF), recommandé par le W3C en 1999, vise à munir le Web d'un modèle de données plus adapté, ayant une structure de *graphe*.
- L'objectif est de construire le *Semantic Web* : un web dans lequel les machines ont (enfin) accès à la sémantique des données.
- Recommandation un peu hâtive, présentant quelques défauts importants (notamment l'absence de sémantique formelle).  
→ faible adoption de RDF



# 1. LOD : de RDF à RDF

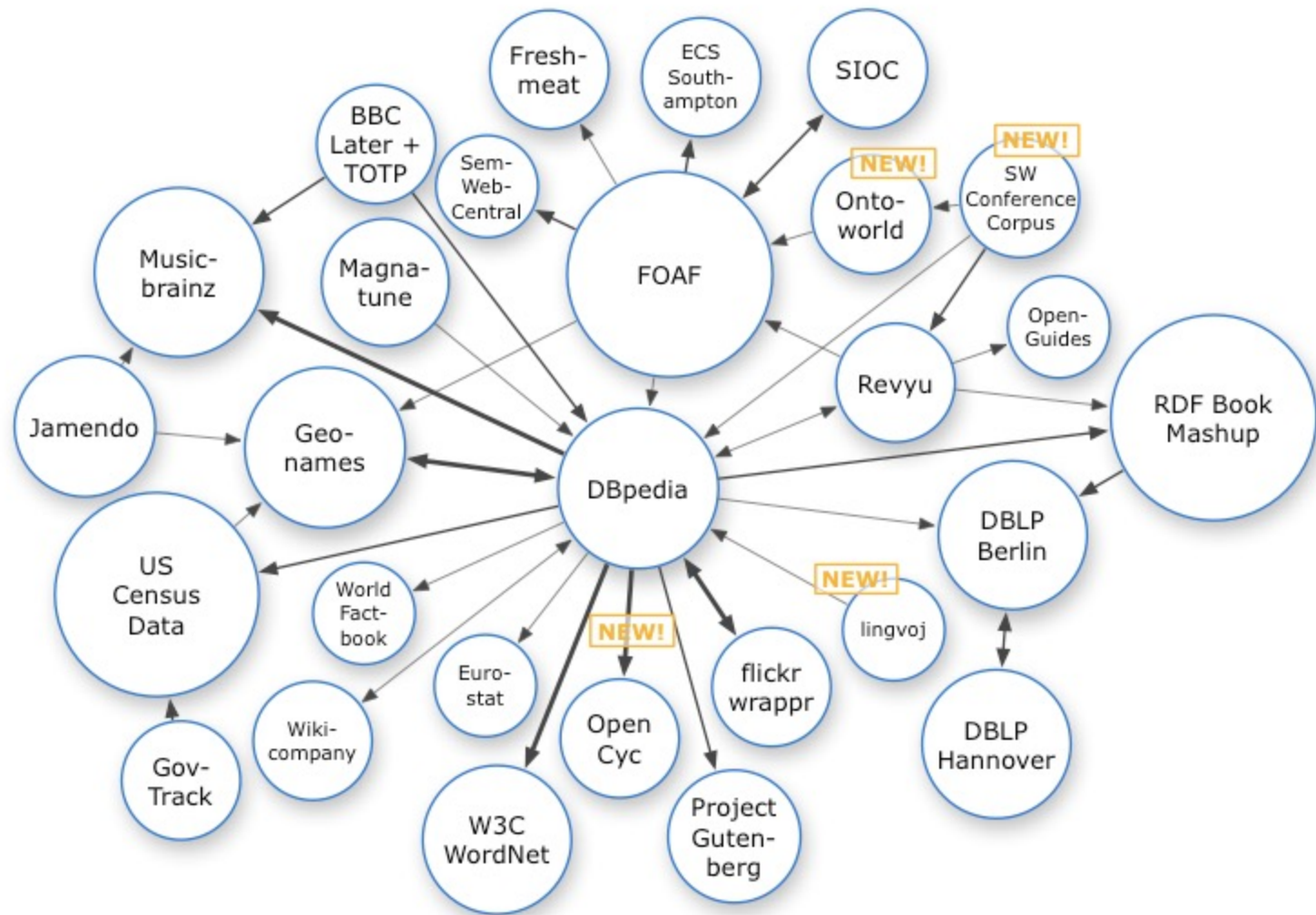
- En 2004, le W3C publie un nouvel ensemble de recommandations sur RDF pour remplacer celles de 1999.
- Pour des raisons de compatibilité avec l'existant, certains aspects sont conservés malgré les débats qu'ils suscitent, mais les défauts considérés comme majeurs sont corrigés.
- Après cet échec relatif, l'appellation *Semantic Web* tombe peu à peu en disgrâce. Certains défenseurs de RDF parlent plus modestement de *Data Web*, puis de *Web of Linked Data* (2006).
- En 2014, RDF est plus largement accepté. Le W3C publie une version révisée (RDF 1.1), endossant notamment plusieurs syntaxes concrètes.

# 1. LOD : Raw Data Now!

- The next Web,  
[https://www.ted.com/talks/tim\\_berners\\_lee\\_the\\_next\\_web](https://www.ted.com/talks/tim_berners_lee_the_next_web)



# 1. LOD : Linked open Data



Source image : [Richard Cyganiak \(https://lod-cloud.net\)](https://lod-cloud.net)

# 1. LOD : Linked open Data

## Les quatre principes de Linked Data

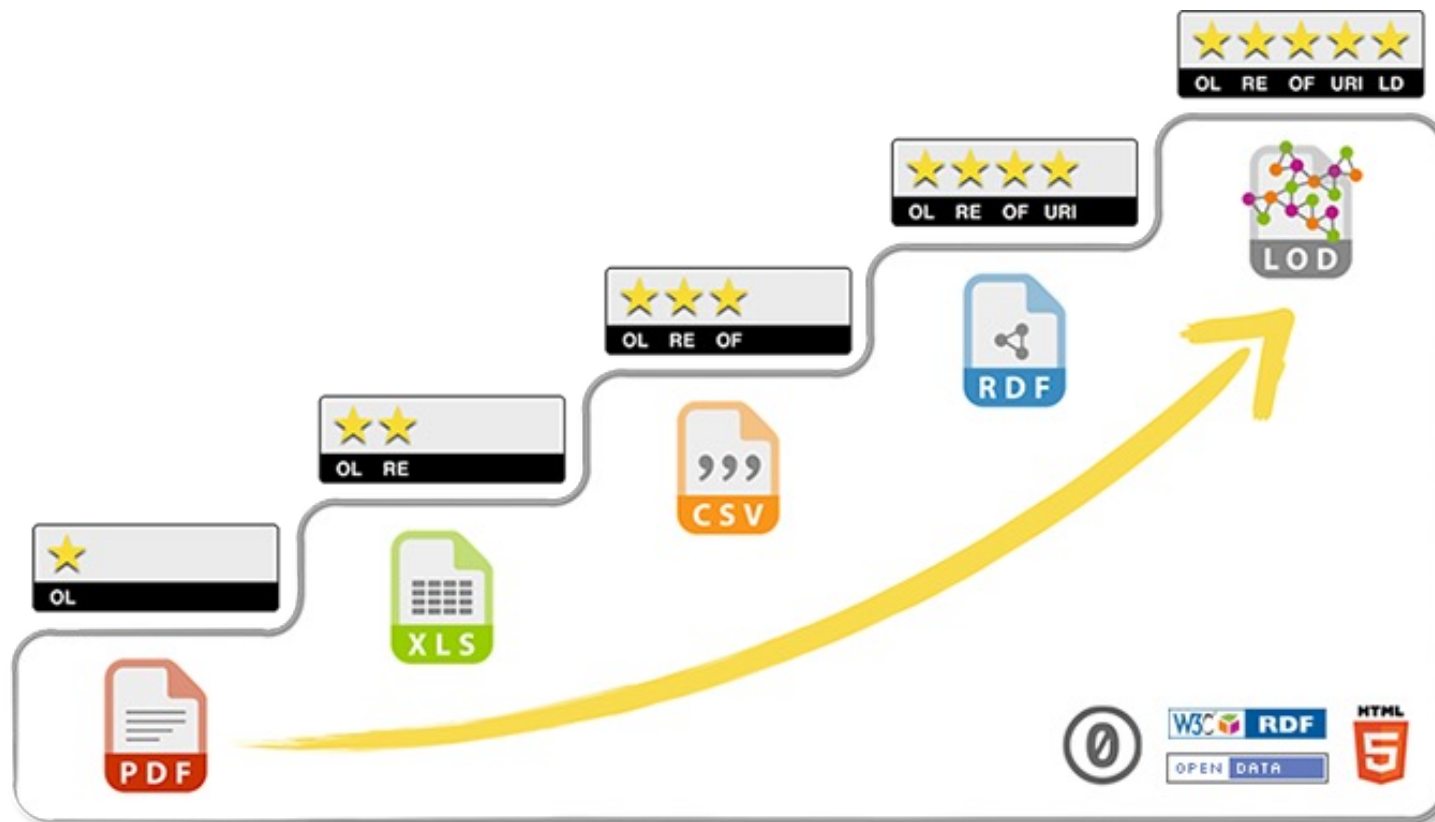
1. Utiliser des IRIs pour nommer les choses (= ressources).
2. Utiliser des IRIs HTTP pour pouvoir obtenir des *représentations* de ces ressources.
3. Fournir ces représentations en utilisant des langages et des protocoles standards ([RDF](#), [SPARQL](#)).
4. Inclure des liens pour permettre de découvrir de nouvelles ressources.

d'après Tim Berners Lee,

<http://www.w3.org/DesignIssues/LinkedData.html>

# 1. LOD : Linked open Data

Linked open data star scheme (<https://5stardata.info/en/>)



# 1. LOD : Projet emblématique, DBPédia

- Projet lancé par Chris Bizer en 2007.
- Objectif : extraire les informations structurées (*infobox*) présentes dans Wikipedia pour les exposer en RDF.
- En novembre 2015 :

The English version of the DBpedia knowledge base describes 4.58 million things, (...) including 1,445,000 persons, 735,000 places (including 478,000 populated places), 411,000 creative works (including 123,000 music albums, 87,000 films and 19,000 video games), 241,000 organizations (including 58,000 companies and 49,000 educational institutions), 251,000 species and 6,000 diseases.

# 1. Projet emblématique : DBPédia

Wikipedia Forever Our shared knowledge. Our shared treasure. Help us protect it. [Show]

## Lyon

From Wikipedia, the free encyclopedia

*This article is about the French city. For other uses, see Lyon (disambiguation).*

**Lyon** (Arpitan: *Liyon*), often Anglicized as **Lyons**, is a city in east-central France in the region Rhône-Alpes, situated between Paris and Marseille. Its name is pronounced [liːz] in French and Arpitan, and /liːˈbɪn or /ˈlɪ.ənz/ in English. Lyon is located at 470 km from Paris, 320 km from Marseille, 160 km from Geneva, 280 km from Turin, 450 km from Milan and 600 km from Barcelona. The residents of the city are called *Lyonnais* (singular and plural).

Lyon is a major center of business with a reputation as the French capital of *gastronomy* and having a significant role in the history of cinema due to Auguste and Louis Lumière. The local professional football team, Olympique Lyonnais, has increased the profile of Lyon internationally through participation in European football championships.


Together with its suburbs and satellite towns, Lyon forms the second-largest metropolitan area in France after that of Paris, with the population of its urban area estimated to be 1,783,400 and that of its metropolitan area 4,415,000<sup>[1]</sup> (2007). Its urban region (*Urban Community of Lyon*), represents half of the Rhône-Alpes *région* population with three million inhabitants<sup>[2]</sup>. Lyon is also a major industrial center specializing in chemical, pharmaceutical, and biotech industries. The city also contains a significant software industry with a particular focus on video games.

Lyon is the *préfecture* (capital) of the Rhône *département*, and also the capital of the Rhône-Alpes *région*. The city is known for its historical and architectural landmarks and is a UNESCO World Heritage Site. Lyon was historically known as the *silk capital of the world*. The city is known as the culinary capital of France. It also hosts the international headquarters of Interpol, Euronews, International Agency for Research on Cancer, and the Cité Scolaire Internationale de Lyon. Lyon is a pilot city of the Council of Europe and the European Commission Intercultural cities programme.

### Contents [hide]

- History
- Geography
  - Climate
  - Administration
  - Arrondissements
- Culture
  - UNESCO World Heritage Site
  - Gastronomy
- Economy
  - Headquarters
  - Business quarter
  - Tourism
  - Research
    - Laboratory P4 Jean Merieux
- Main sights
  - Noteworthy sites
  - Parks and gardens
- Colleges and universities
  - International primary and secondary level schools in Lyon
- Sport
- Transport
  - Air
  - Rail
  - Road
  - Intercity coach
  - Public transport
  - International attraction



### Location




Time zone: CET (GMT+1)

Administration	
Country	France
Region	Rhône-Alpes
Department	Rhône (69)
Arrondissement	Lyon
Canton	chief town of 14 cantons
Subdivisions	9 arrondissements
Intercommunality	Urban Community of Lyon

### Ville de Lyon



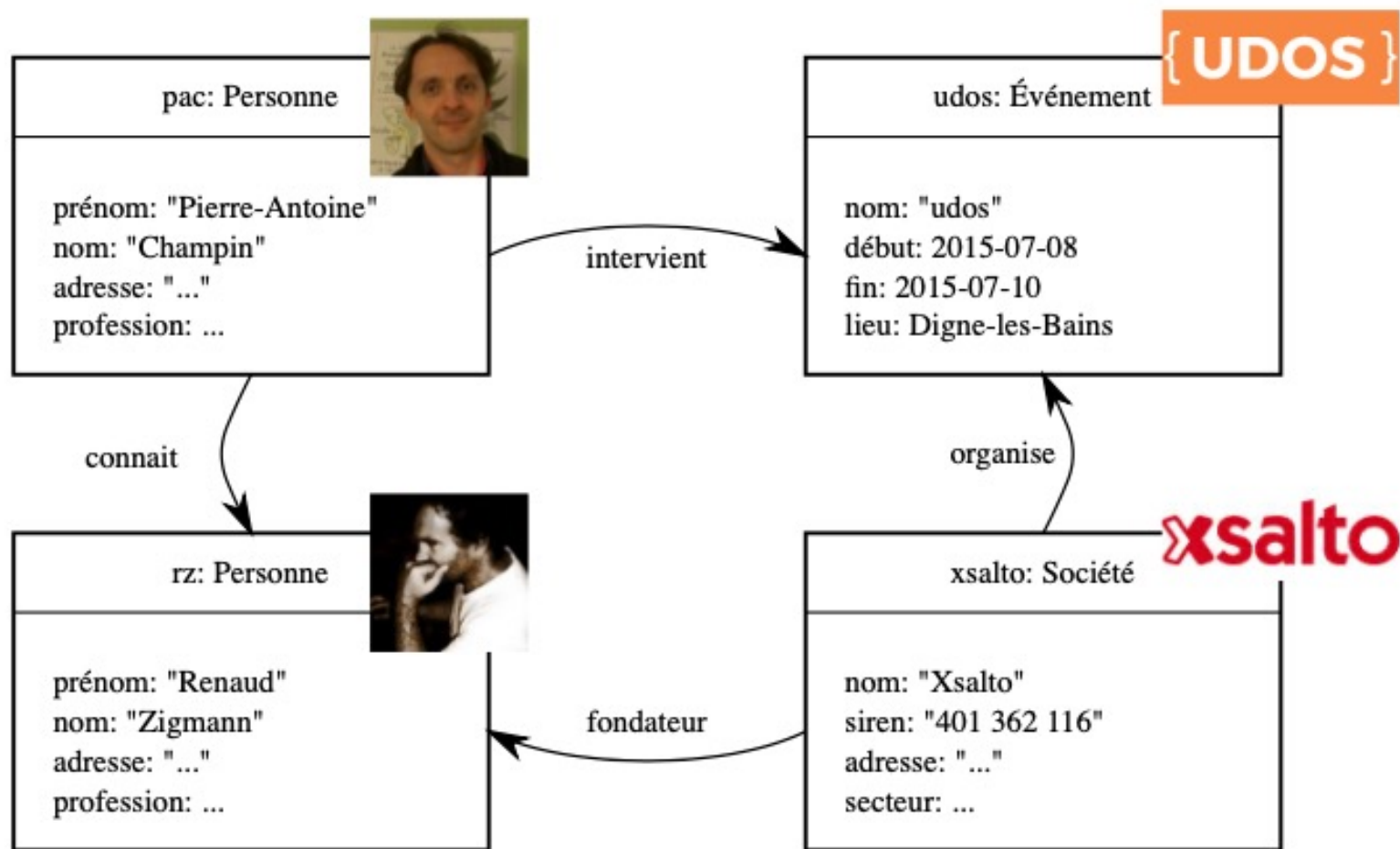
Motto: *Avant, avant, Lion le meilleur*  
(Arpitan: *Forward, forward, Lyon the best*)



Lyon as seen from Fourvière

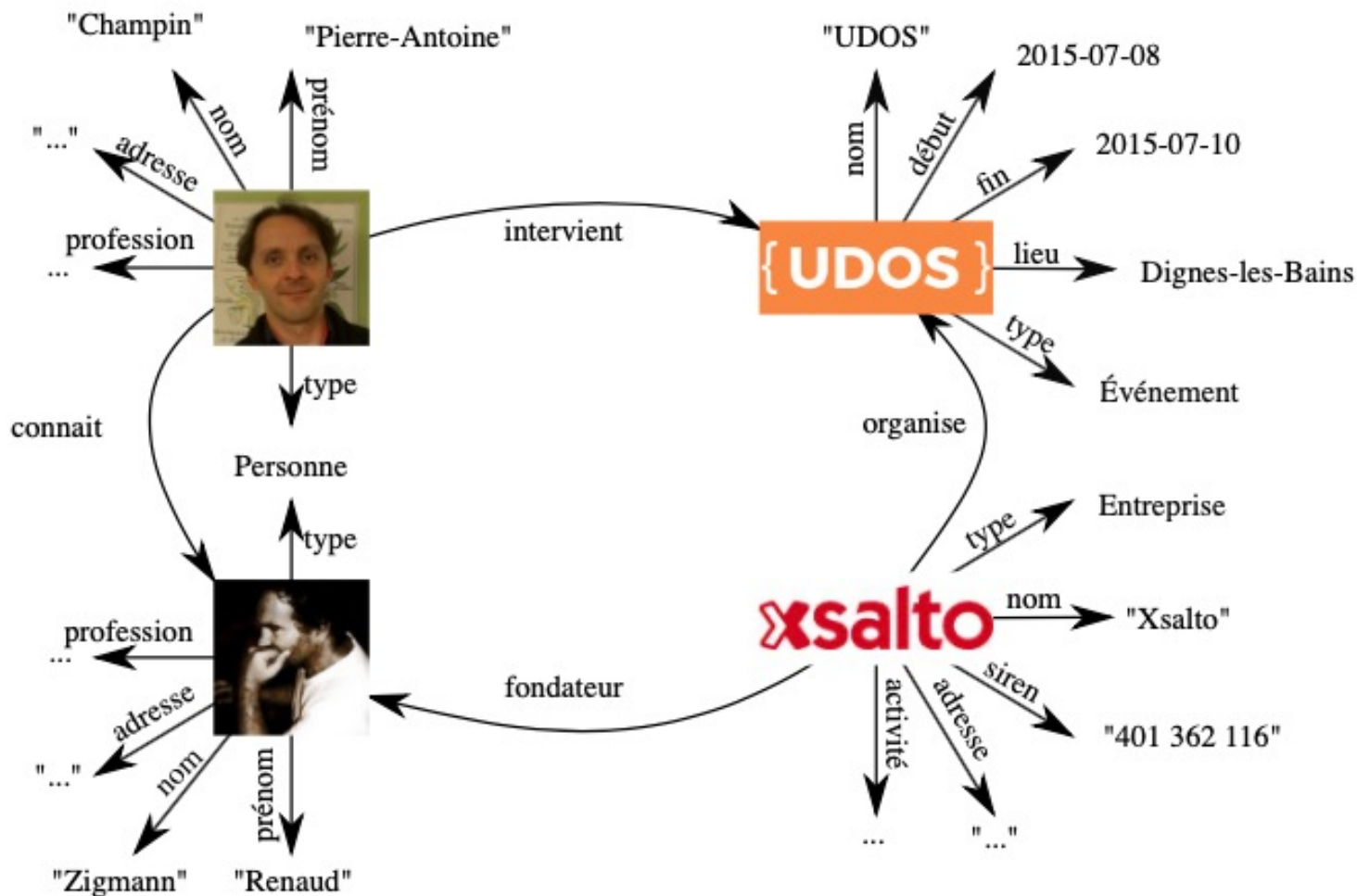


## 2. RDF : des données aux données liées



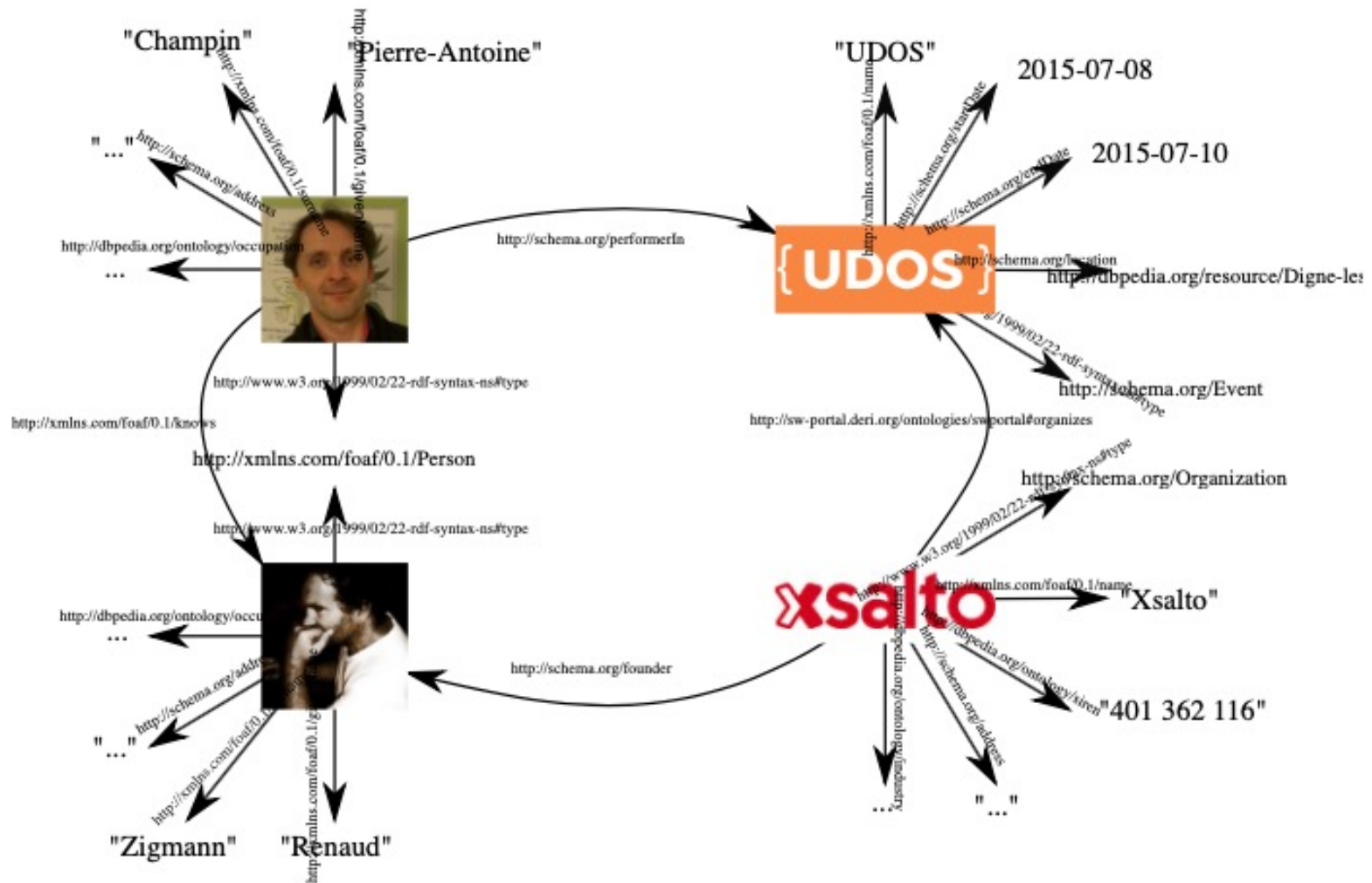
Données classiques

## 2. RDF : des données aux données liées



Données vues comme un graphe

# 2. RDF : Des données aux données liées



Données vues comme un graphe avec des IRIs